How do CSD turn the promise of Big Data and advance analytics into value?

The 21st ACG Cross Training Seminar

Yogyakarta, 4th – 7th March 2019

Big data and advanced analytics

KSEI big data initiatives

![ksei - Indonesia Central Securities Depository]
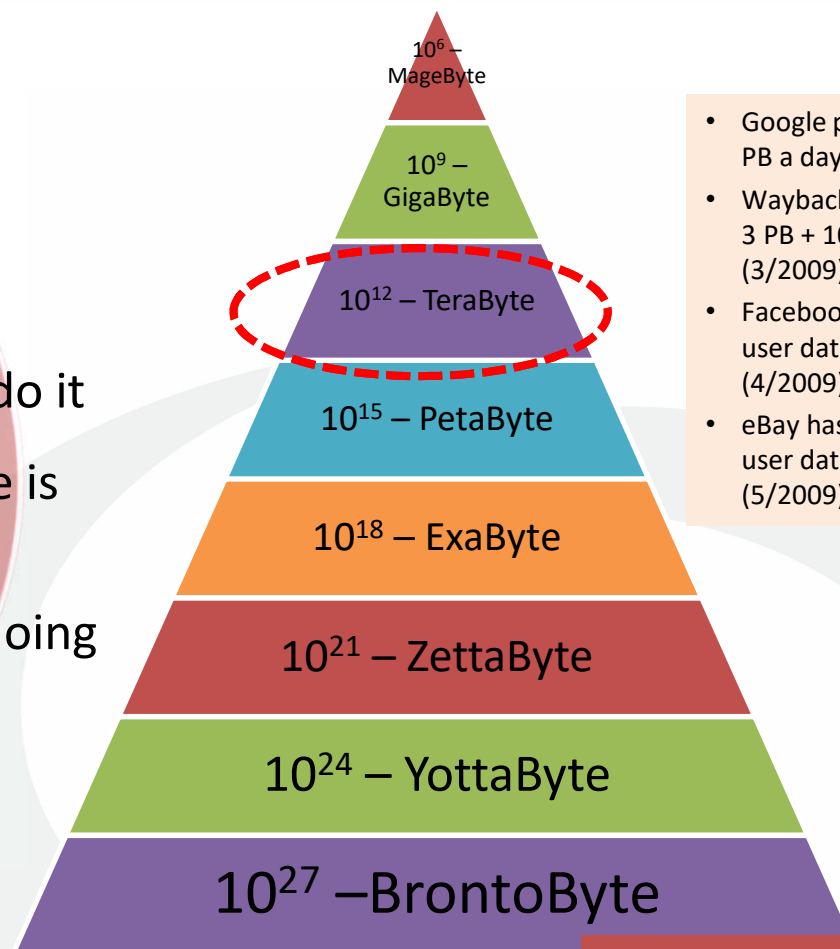
Big Data

Everyone talks about it

Nobody really knows how to do it

Everyone thinks everyone else is doing it

So everyone claims they are doing it

Pyramid (top to bottom):
- 10^6 – MageByte
- 10^9 – GigaByte
- 10^{12} – TeraByte
- 10^{15} – PetaByte
- 10^{18} – ExaByte
- 10^{21} – ZettaByte
- 10^{24} – YottaByte
- 10^{27} – BrontoByte

- Google processes 20 PB a day (2008)
- Wayback Machine has 3 PB + 100 TB/month (3/2009)
- Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)

The digital universe of tomorrow

Big data happens when the data you have to process is bigger than what you can process in the given time with the current technology

Big data is when we can handle data fast enough to make a difference

What is big data

The basic idea behind phrase Big Data is that everything we do is increasingly leaving a digital trace (or data), which we (and other) can use and analyse.
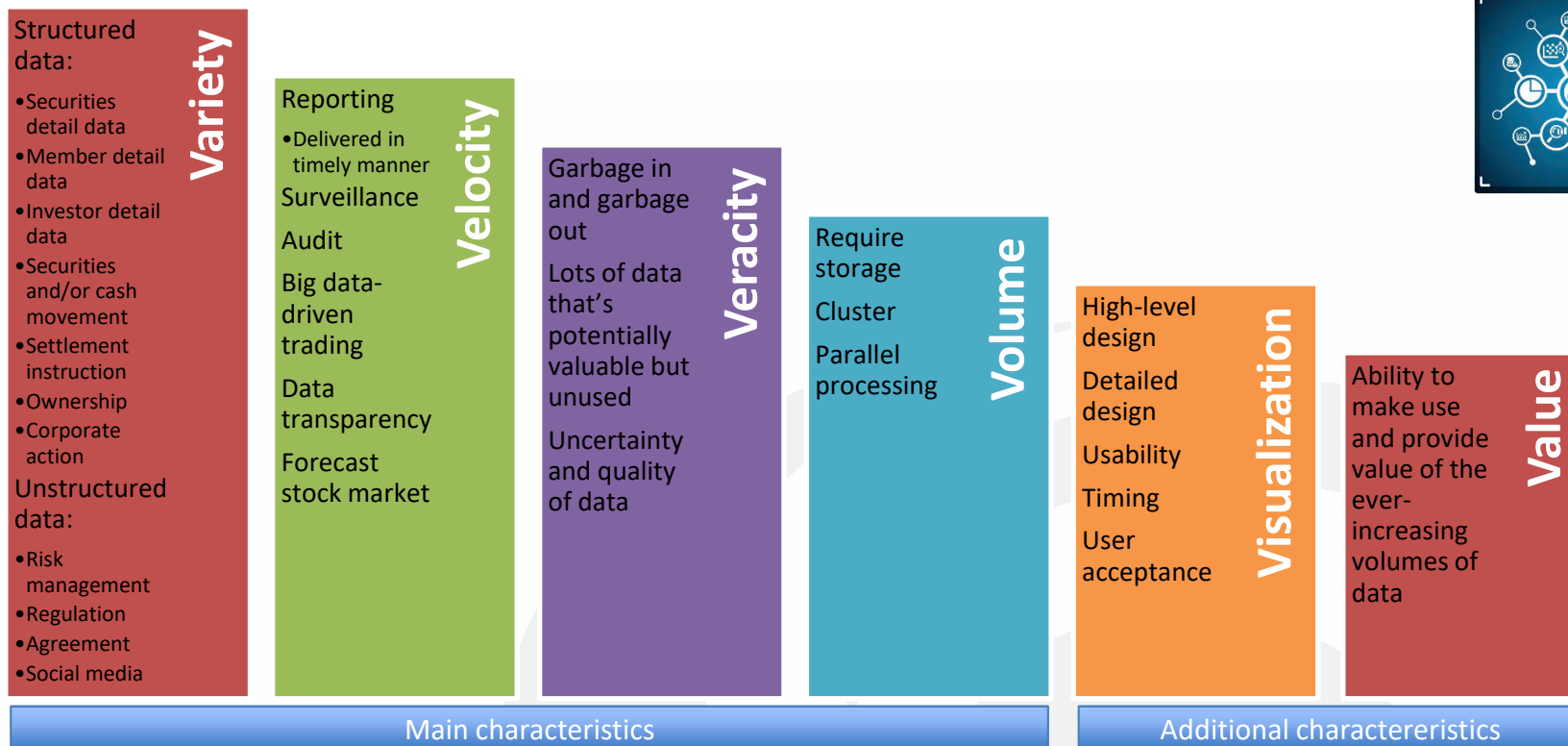
Big data is similar to small data, but bigger in size

But having bigger, it requires different approach: technique, tools, and architecture

Big data generates values from the storage and processing of every large quantities of digital information that cannot be analysed with traditional computing techniques.
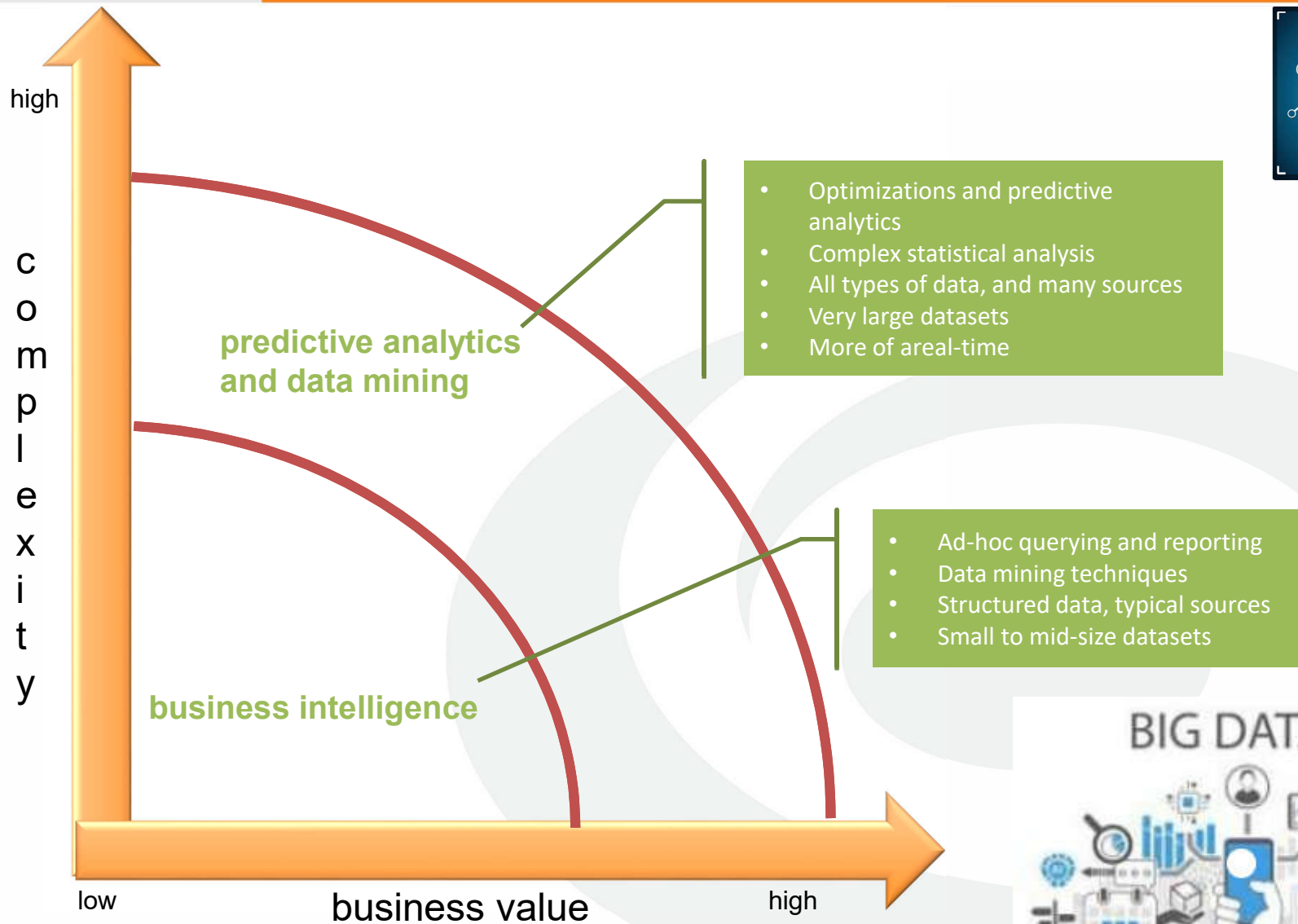
An aim to solve new problem or old problem in a better way

Big data characteristics

Variety

Structured data:
- Securities detail data
- Member detail data
- Investor detail data
- Securities and/or cash movement
- Settlement instruction
- Ownership
- Corporate action

Unstructured data:
- Risk management
- Regulation
- Agreement
- Social media

Velocity

Reporting
- Delivered in timely manner

Surveillance

Audit

Big data-driven trading

Data transparency

Forecast stock market

Veracity

Garbage in and garbage out

Lots of data that's potentially valuable but unused

Uncertainty and quality of data

Volume

Require storage

Cluster

Parallel processing

Visualization

High-level design

Detailed design

Usability

Timing

User acceptance

Value

Ability to make use and provide value of the ever-increasing volumes of data

Main characteristics

Additional charactereristics

Turning Big Data into Value

The datafication of our world give us unprecedented amounts of data in term of volume, velocity, variety, veracity, and visualization. The latest technology such as cloud computing and distributed systems together with the latest software and analysis approaches allow us to leverage all types of data to gain insights and add value.

What's driving big data in CSD

high

complexity

predictive analytics and data mining

- Optimizations and predictive analytics
- Complex statistical analysis
- All types of data, and many sources
- Very large datasets
- More of areal-time

- Ad-hoc querying and reporting
- Data mining techniques
- Structured data, typical sources
- Small to mid-size datasets

business intelligence

low   business value   high

BIG DATA

Big data and advanced analytics

The big data analytics improved in the recent digital world as real time data had been driving the business

The era of **big data** is evolving rapidly,

Time has come to define a pragmatic approach to **big data** and **advanced analytics** -

and experiences suggest that **CSD** should act now.

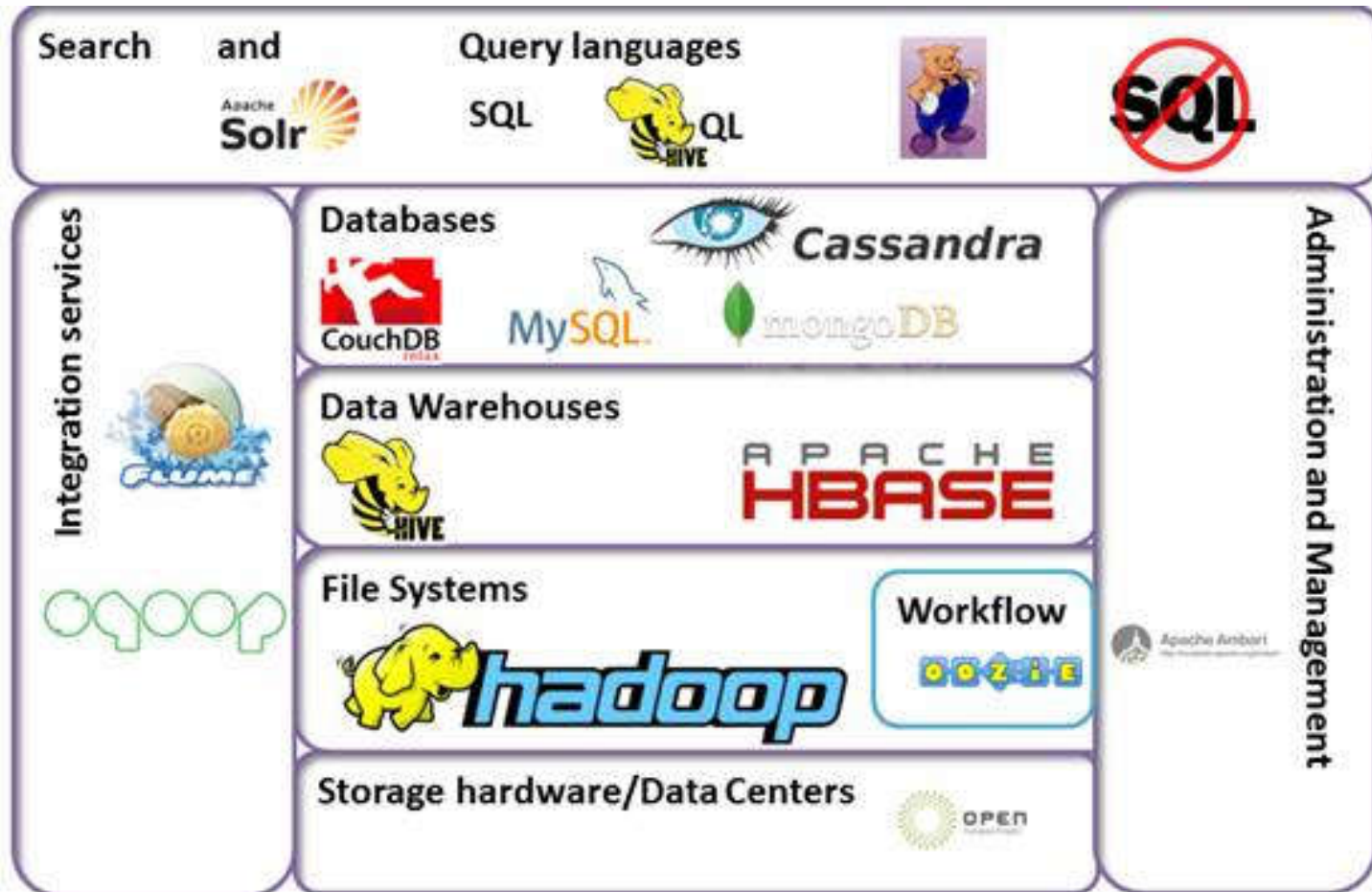*tightly focused on how to use data to make better decisions.*

Big data and analytics solution

A simple effective SWAT framework to help business leaders translate big data into actionable intelligence (Carter, 2014).

- Question
- Business case
- Metrics/levers
- Rout cause

Strategic business question

Wrangle data

- Business ownership
- Data definition
- Source
- Analysis
- Transform
- Validation
- Timing
- Availability
- Documentation

Take action

Answer with visualization

- Process change
- Organization change
- Value realization
- Education/training
- More questions
- Iterate

- High level design
- Detailed design
- Usability
- Timing
- Build tool
- Process change
- User acceptance

Source:
Carter, K. B. (2014). *Actionable Intelligence.* New Jersey: Jhon Wiley & Sons, Inc.

Benefit big data and advanced analytics for CSD

Real time big data isn't just a process for storing petabytes or exabytes of data in a data warehouse. It's about the ability of CSD to make a better decision and take meaningful action in the right time.
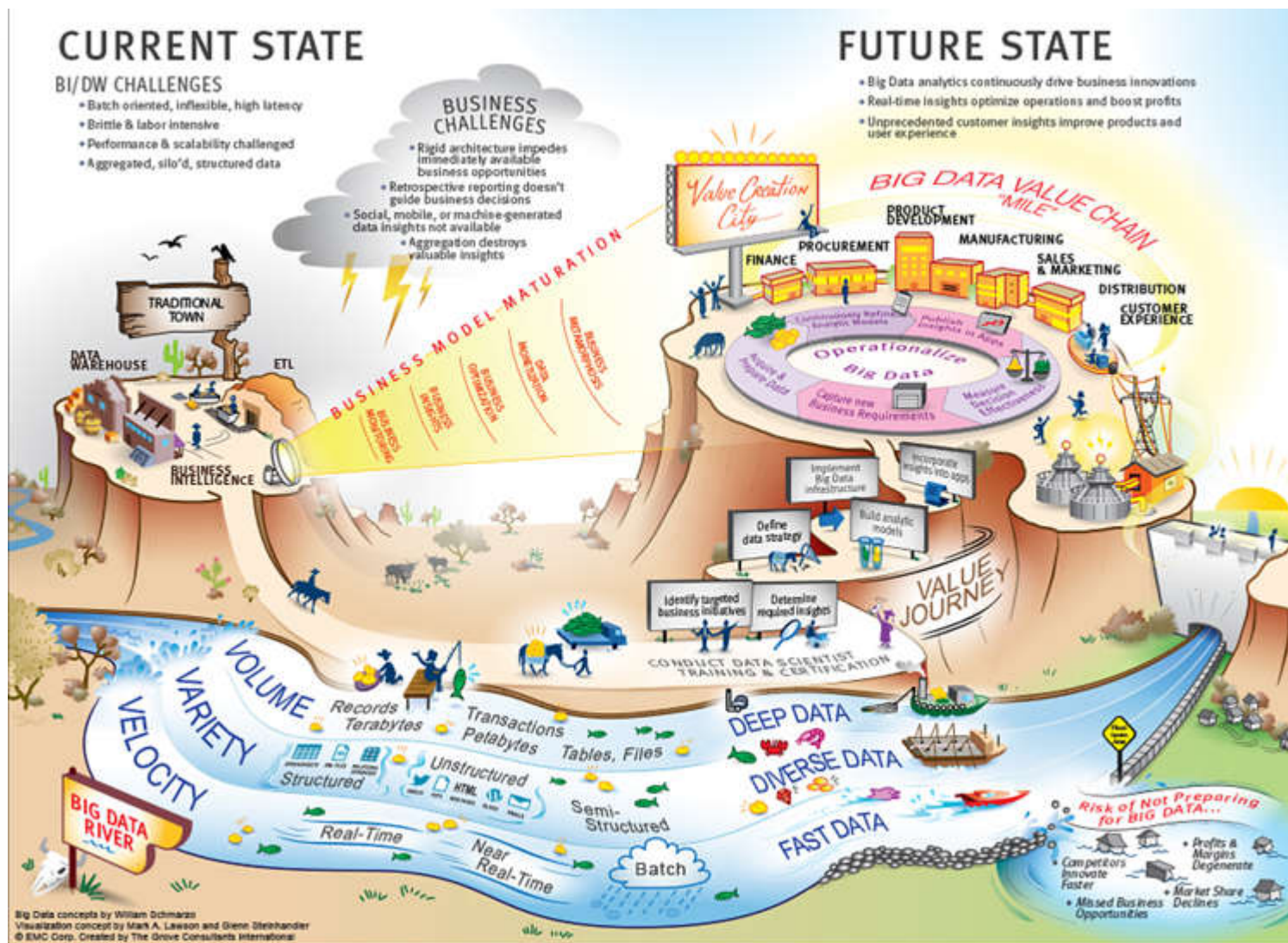
Fast forward to the present and technologies like Hadoop give CSD a scale and flexibility to store data before CSD know how are CSD going to process it.

Technologies such as MapReduces, Hive, and Impala enable CSD to run without changing the data structures underneath.

The CSD are using big data to target customer-centric (participant or investors) outcomes, tap into internal data and build a better information ecosystem.

It offer commercial opportunities for CSD of a comparable scale to enterprise software in the late 1980s and the internet boom of the 1990s, and the social media explosion of today
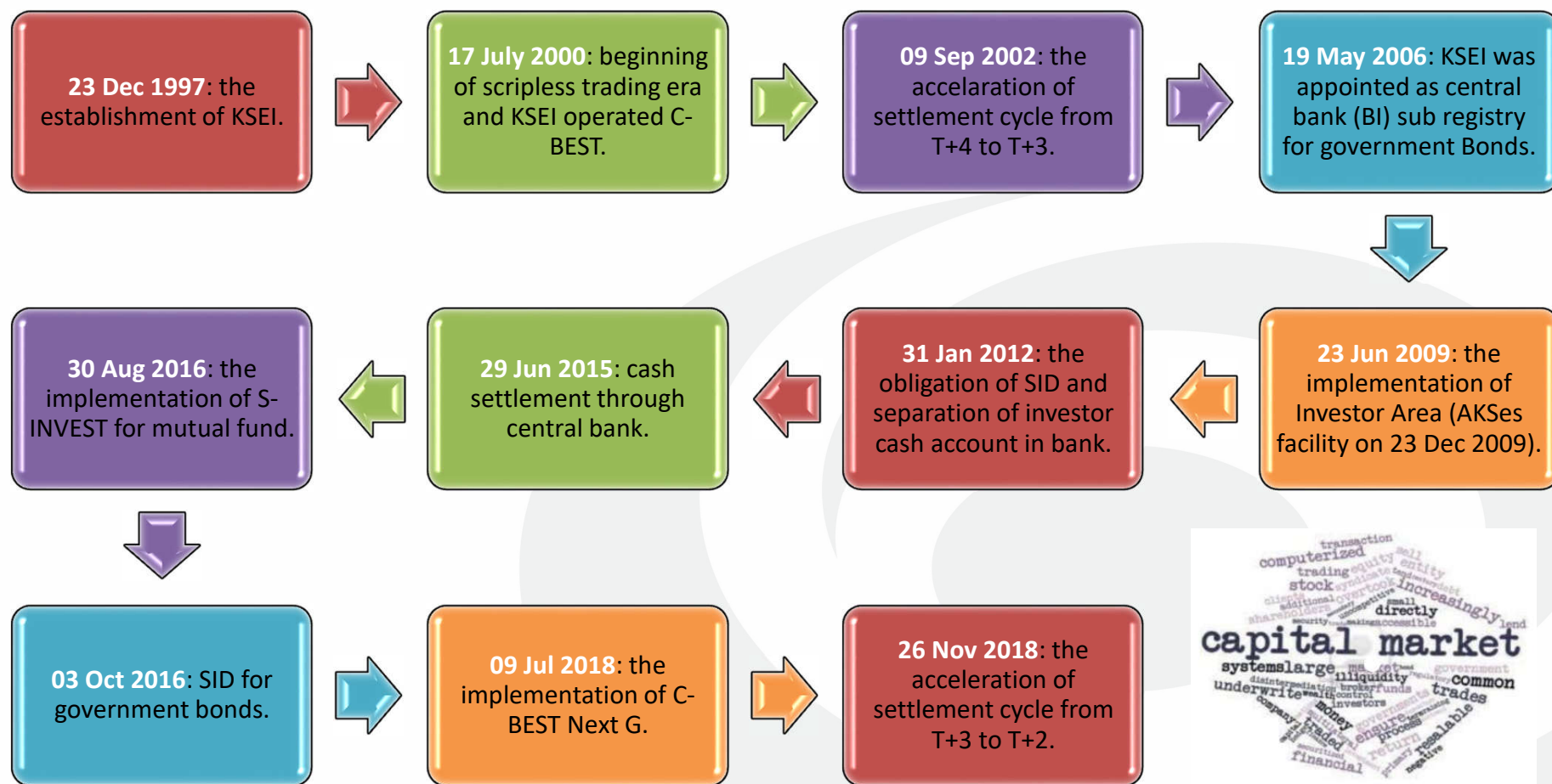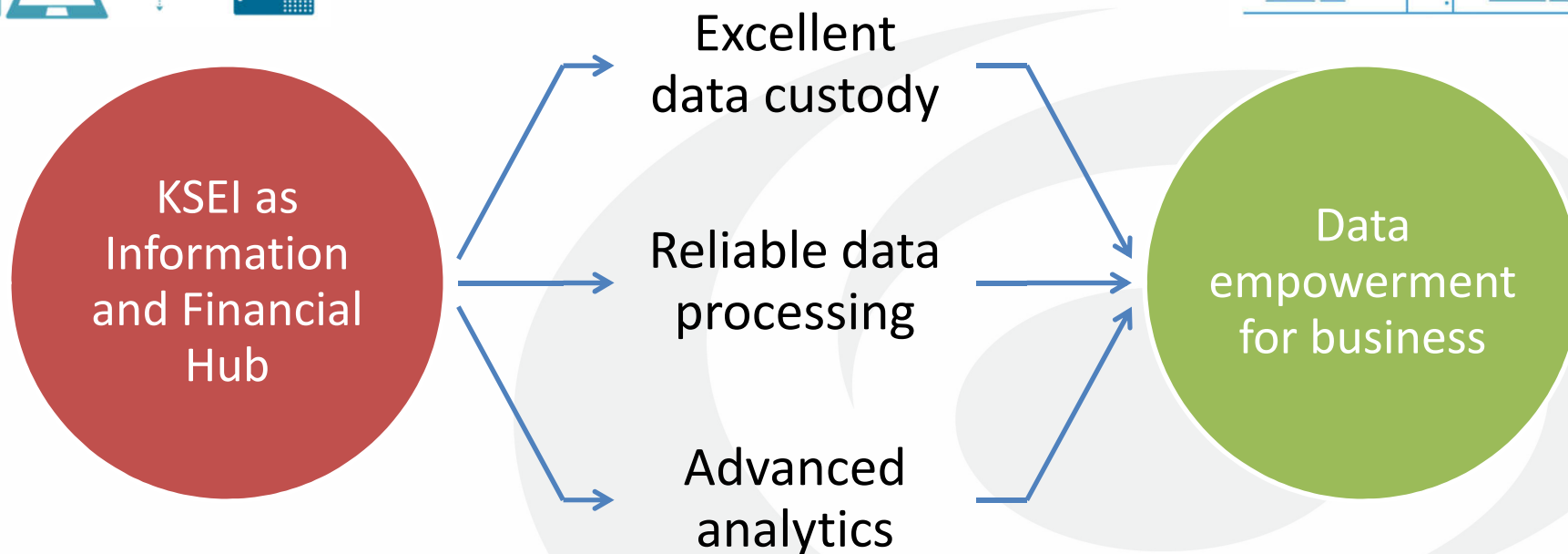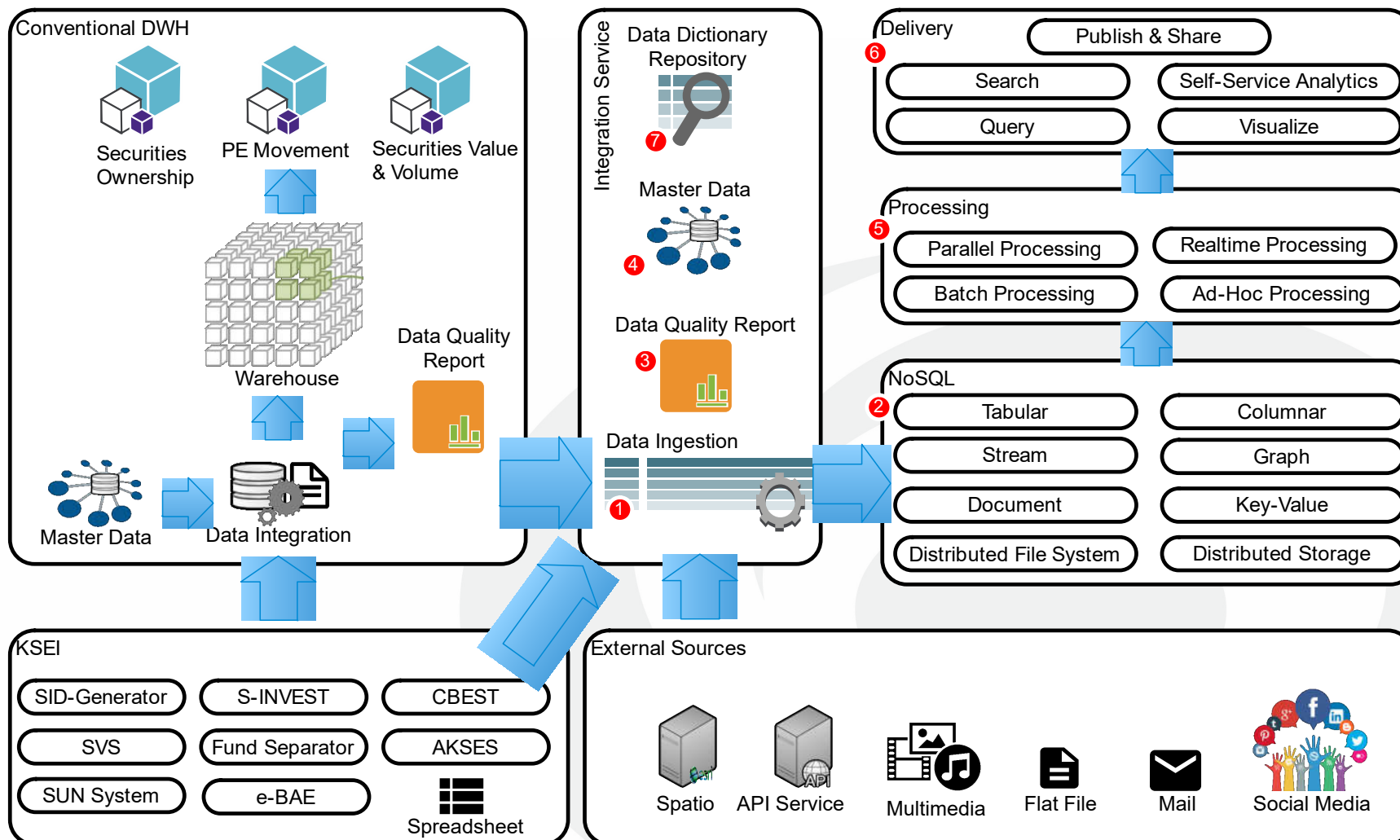
Big data story map

Big data and advanced analytics

KSEI big data initiatives

KSEI journey

![ksei logo] Indonesia Central Securities Depository

23 Dec 1997: the establishment of KSEI.

→

17 July 2000: beginning of scripless trading era and KSEI operated C-BEST.

→

09 Sep 2002: the acceleration of settlement cycle from T+4 to T+3.

→

19 May 2006: KSEI was appointed as central bank (BI) sub registry for government Bonds.

↓

30 Aug 2016: the implementation of S-INVEST for mutual fund.

←

29 Jun 2015: cash settlement through central bank.

←

31 Jan 2012: the obligation of SID and separation of investor cash account in bank.

←

23 Jun 2009: the implementation of Investor Area (AKSes facility on 23 Dec 2009).

↓

03 Oct 2016: SID for government bonds.

→

09 Jul 2018: the implementation of C-BEST Next G.

→

26 Nov 2018: the acceleration of settlement cycle from T+3 to T+2.

KSEI big data initiatives

KSEI as Information and Financial Hub

→ Excellent data custody

→ Reliable data processing

→ Advanced analytics

Data empowerment for business

KSEI big data to be

Conventional DWH

Securities Ownership

PE Movement

Securities Value & Volume

Warehouse

Data Quality Report

Master Data

Data Integration

Integration Service

Data Dictionary Repository ⑦

Master Data ④

Data Quality Report ③

Data Ingestion ①

Delivery ⑥

Publish & Share

Search

Self-Service Analytics

Query

Visualize

Processing ⑤

Parallel Processing

Realtime Processing

Batch Processing

Ad-Hoc Processing

NoSQL ②

Tabular

Columnar

Stream

Graph

Document

Key-Value

Distributed File System

Distributed Storage

KSEI

SID-Generator

S-INVEST

CBEST

SVS

Fund Separator

AKSES

SUN System

e-BAE

Spreadsheet

External Sources

Spatio

API Service

Multimedia

Flat File

Mail

Social Media

Counterpart data, social media, voice, news
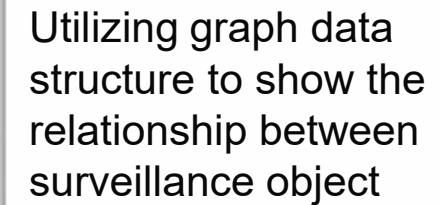
Data warehouse

- Investor
- Securities ownership
- Transaction
- Movement
- Mutual fund activities
- Scrip Securities ownership

Strategic use case

- Surveillance
- Potential fraud identification
- Operation efficiency
- Trend prediction
- Risk controlling
- etc
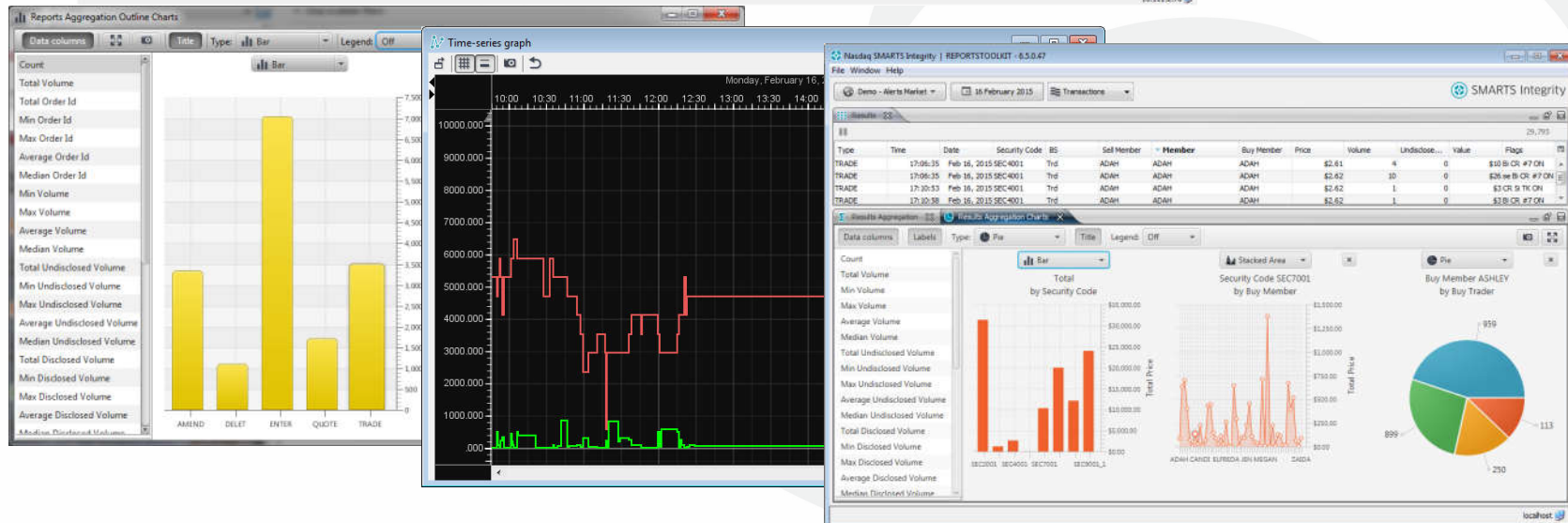
KSEI's SMARTS surveillance system

Utilizing graph data structure to show the relationship between surveillance object
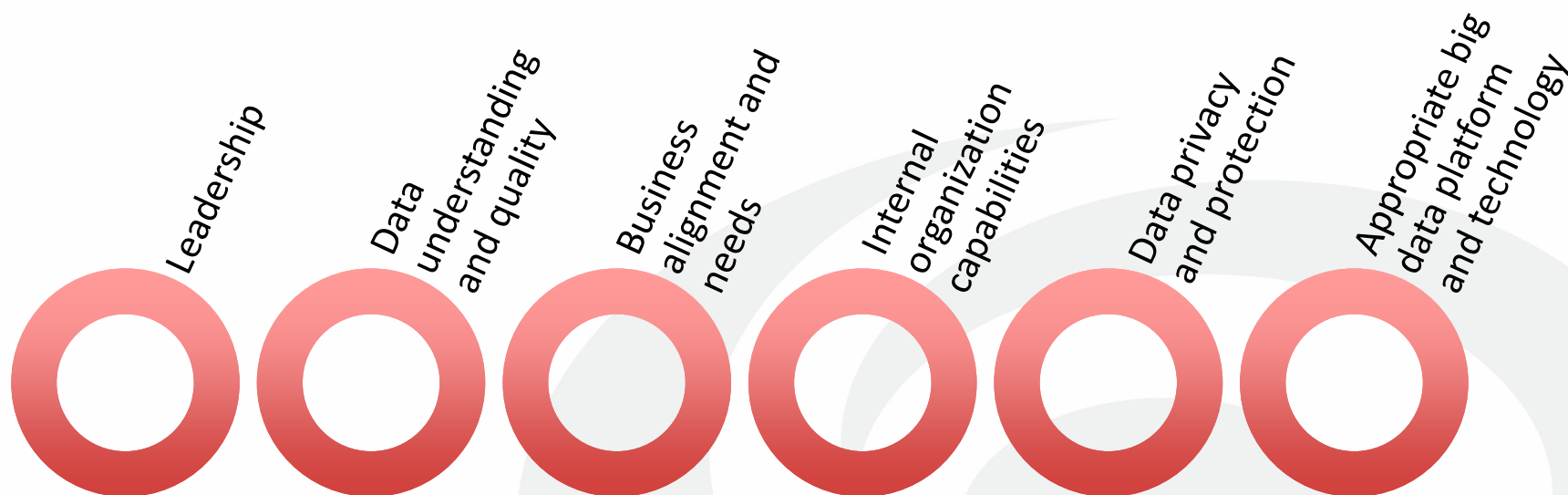
Thank You
Arigato Gozaimasu
Merci
Shukraan
Xièxiè
Terima kasih